TITLE OF THE INVENTION

ASSIGNMENT OF PHONEMES TO THE GRAPHEMES PRODUCING THEM

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is based on and hereby claims priority to German Application No. 10042943.2 filed on August 31, 2000 in Germany, the contents of which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

[0002] The invention relates to a method, a computer program product, a data medium and a computer system for the assignment of phonemes to the graphemes producing them in a lexicon having words (grapheme sequences) and their associated phonetic transcription (phoneme sequences).

[0003] Speech processing methods are disclosed, for example, in US 6,029,135, US 5,732,388, DE 19636739 C1 and DE 19719381 C1. Routines for grapheme-phoneme conversion, that is to say for converting written words into spoken sounds, are required for automatically reading aloud or extending the vocabulary of dictation systems or of automatic speech recognition systems. Neural networks are frequently used for this purpose.

[0004] The training of these neural networks is performed with the aid of patterns. A pattern includes of a number of letters from a word which are applied to the input nodes of a neural network, and of the associated phoneme corresponding to the output node. Each phoneme is frequently also assigned what is termed a grouping value. The grouping value specifies the number of graphemes which produce the associated phoneme.

[0005] The patterns are obtained from what are termed training lexica. A training lexicon contains assignments of graphemes, as a rule words, numerals, etc., that is to say everything which is to be converted, to phonemes and phoneme sequences, that is to say grapheme-phoneme transcriptions at the level of words. The phoneme sequences are produced in the training lexicon by a suitable type of phonetic transcription. SAMPA phonetic transcriptions or Spicos inventory, which are based on ASCII characters, are frequently used in the field of automatic speech recognition. A few German words may be listed by way of example with the associated phonetic transcription in SAMPA:

Quatsch	kv'atS
spät	SpE:t
Schutz	SUts
schwer	Sve:6
Sprache	Spra:x@

[0006] The sound "sch" is represented, for example, by [S], lengthenings by a colon. In this case, phonemes are represented in square brackets [], graphemes in pointed brackets <>. All the examples of phonetic transcription in the description are reproduced in SAMPA.

[0007] Although these training lexica include the phonetic transcription, they do not include the unique assignment of phonemes and the graphemes producing them, as required for the patterns. For example, the following assignment would be desirable for the word <Sprache>:

Graphemes	S	p	r	a	c	h	e
Phonemes	S, 1	p, 1	r, 1	a:, 1	x, 2		@, 1

from which it is easier to derive the patterns for training the neural network. In the case of an input window with 7 letters, the following 6 patterns are yielded directly from the unique assignment:

1st Pattern	Input				S	p	r	a
	Output	S, 1						

The grapheme sequence of 3 empty characters, <S>, <p>, <r> and <a>, <S> being located centrally in the input window, is assigned to the sound [S] with the grouping value 1. The following are obtained correspondingly as further patterns:

2nd Pattern	Input			S	p	r	a	c
	Output	p, 1						

3rd Pattern	Input		S	p	r	a	c	h
	Output	r, 1						

4th Pattern	Input	S	p	r	a	c	h	e
	Output	a:, 1						

5th Pattern	Input	p	r	a	c	h	e	
	Output	x, 2						

The "Ach" sound, or voiceless velar fricative "ch" is assigned a grouping value of 2 in accordance with the segmentation rules, since it is assigned the two letters <c> and <h>. The letter window can therefore be displaced in the following pattern by 2 letters:

6th Pattern	Input	a	c	h	e			
	Output	@, 1						

[0008] The assignment of letters to phonemes is not, however, yielded uniquely from the phonetic transcription of the lexicon. The word <Sprache> has of 7 letters, but only of 6

phonemes. The question arises as to which of the phonemes is produced by 2 letters. Since also 2 phonemes can be produced by one letter, for example [ks] by <x>, the uncertainty in the grapheme-phoneme assignment is a general problem for the patterns.

[0009] To date, the grapheme-phoneme assignment has been carried out semi-automatically, starting from empirical rules evident to a native speaker, but this is subject to error, particularly in the case of multilingual systems, and constitutes a substantial outlay.

SUMMARY OF THE INVENTION

[0010] It is an object of one aspect of the invention automatically to produce the assignment of phonemes to the graphemes producing them for patterns for training a neural network for grapheme-phoneme conversion.

[0011] In this case, in the context of a computer program product the computer program is understood as a suitable product in whatever form, for example on paper, on a machine-readable data medium, distributed over a network, etc.

[0012] According to one aspect of the invention, the assignment of phonemes to the graphemes producing them is carried out in a lexicon having words (grapheme sequences) and their associated phonetic transcription (phoneme sequences) with the aid of a dynamic time warping (DTW) algorithm.

[0013] DTW algorithms are a variant of dynamic programming. They are described, for example, in:

1. Hoffmann, R.: "Signalanalyse und -erkennung" (Signal analysis and recognition.), Springer Verlag, Berlin, Heidelberg, 1998, pages 390-393.

2. Rabiner, L.R.; Juang, B.-H.: "Fundamentals of speech recognition." Englewood Cliffs: Prentice Hall 1993 (Prentice Hall Signal Processing Series).

3. Besling, S.: "Heuristical and Statistical methods of Grapheme-to-Phoneme Conversion"; Proceedings KONVENS 94, Vienna, pages 23-31.

[0014] It is preferred to select in a first step words in which the number of the graphemes and the number of the phonemes coincide. In these words, the graphemes and phonemes are assigned to one another in the sequence of the specification of their graphemes and phonemes in the lexicon. The relative frequency with which a phoneme is produced by a grapheme is determined from these assignments. Alternatively, it is also possible to determine the relative frequency with which a grapheme is assigned to a phoneme.

[0015] Created in a second step for each word of the lexicon is a two-dimensional matrix, the so-called incidence matrix, one index of which is given by the grapheme of the word, and the second index of which is given by the phoneme of the word. The relative frequencies belonging to the respective phoneme-grapheme pair and determined in the first step are selected as entries of the matrix.

[0016] In a third step, each matrix entry is logically combined by a mathematical operation, in particular a multiplication, with the extreme value, which is preferably the maximum value, of the following three preceding matrix entries: the entry for the same phoneme and the preceding grapheme in the word, the entry for the preceding phoneme and the same grapheme in the word, and the entry for the preceding phoneme and the preceding grapheme in the word. Other computing operations are also conceivable instead of multiplication, for example addition of the reciprocals of the matrix entries, or other operations successful in dynamic programming.

[0017] The first grapheme and the first phoneme of the word are the starting point in the multiplication operation, the modified entries of the matrix respectively yielded from the multiplication operations being used in determining the maximal values. A step direction is determined for this matrix entry by determining which of the three preceding matrix entries was extreme.

[0018] In a fourth step, the step direction determined for the matrix entry is respectively defined, starting from the matrix entry for the last phoneme and the last grapheme, along a path through the matrix up to the matrix entry for the first phoneme and the first grapheme. The matrix elements belonging to the path define the assignment of graphemes to phonemes of the word.

[0019] The lexicon is therefore consistently prepared. The method according to one aspect of the invention can be adapted for producing patterns for training neural networks.

[0020] After execution of the assignment of graphemes to phonemes for each word of the lexicon, these assignments are used to determine the position-dependent relative frequency with which a phoneme is produced by two or more graphemes, or two or more phonemes are produced by a grapheme, or two or more graphemes are assigned to a phoneme, or a grapheme is assigned to two or more phonemes. This permits corrections to be undertaken to the assignments in a further step.

[0021] These corrected assignments can be used for iterative improvements of the relative frequencies and thus of the assignments. For this purpose, after the correction of the assignments, the position-dependent relative frequencies are determined anew for each word of the lexicon from these corrected assignments. These are used in further assignments.

[0022] When determining the relative frequencies, it is advantageous to take into account only those assignments in which the matrix entry for the last phoneme and the last grapheme exceeds a prescribed threshold value after execution of the multiplications. This filters out long words in the case of which the assignment is uncertain, as well as very rare and therefore uncertain assignments.

[0023] It is advantageous to use unique entry knowledge for the matrix entries in order to create stable fixed points. Thus, for example, the matrix entry for the first phoneme and the first grapheme of each word is set to 1, like the matrix entry for the last phoneme and the last grapheme of each word. These two entries form the starting point and finishing point, respectively, of the path to be determined, and must be traversed in any case. On the other hand, the matrix entry for the first phoneme and the last grapheme of each word, as well as the matrix entry for the last phoneme and the first grapheme of each word are set to 0, because these assignments are basically ruled out.

[0024] The diagonal is preferred as the most likely path when determining the maximum in conjunction with the multiplication. That is to say, if in the determination of the maximum value of the three preceding matrix entries the matrix entry for the preceding phoneme and the preceding grapheme in the word and one of the other two entries are of equal magnitude, the matrix entry for the preceding phoneme and the preceding grapheme in the word is regarded as a maximum.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] These and other objects and advantages of the present invention will become more apparent and more readily appreciated from the following description of the preferred embodiments, taken in conjunction with the accompanying drawings of which:

Fig. 1 shows a computer system suitable for assigning phonemes to the graphemes producing them in a lexicon;

Fig. 2 shows a matrix with a 1-to-1 assignment of graphemes and phonemes for the word <haben>;

Fig. 3 shows a matrix for assigning graphemes and phonemes for the word <textlich>;

Fig. 4 shows the matrix of the transition frequencies for the assignment of graphemes and phonemes for the word <können>;

Fig. 5 shows the matrix in accordance with Fig. 4 after execution of multiplications; and

Fig. 6A shows a matrix in accordance with Fig. 5 for the word <yield>; and

Fig. 6B shows the matrix in accordance with Fig. 6A after a correction of the assignment of graphemes and phonemes.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0026] Reference will now be made in detail to the preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings, wherein like reference numerals refer to like elements throughout.

[0027] Fig. 1 shows a computer system suitable for assigning phonemes to the graphemes producing them. This system has a processor (CPU) 20, a main memory (RAM) 21, a program memory (ROM) 22, a hard disk controller (HDC) 23, which controls a hard disk (30), and an interface controller (I/O controller) 24. The processor 20, main memory 21, program memory 32, hard disk controller 23 and interface controller 24 are coupled with one another via a bus, the CPU bus 25, for exchanging data and commands. The computer also has an input/output bus (I/O bus) 26, which couples various input and output devices to the interface controller 24. The input and output devices include, for example, a general input and output interface (I/O interface) 27, a display 28, a keyboard 29 and a mouse 31.

[0028] It is described below how the assignment of phonemes to graphemes producing them is carried out for a word.

[0029] Various relative frequencies for calculating the best assignment are used in the following description, and are generally denoted below briefly as frequencies. The frequency with which the grapheme g is assigned to the phoneme p is also termed the transitional frequency and is calculated from

$$H(g \rightarrow p) = \frac{Z(g \rightarrow p)}{N(p)}$$

[0030] In this case, Z(g->p) is the number of assignments of the grapheme g, denoted below by <g>, the phoneme p, denoted below by [p], and N(p) is the number of all the assignments of all the graphemes to this phoneme [p].

[0031] Further frequencies are also required, since the relative frequency of the direct assignment of a grapheme to a phoneme is not sufficient for a final decision on the assignments. Consequently, position-dependent frequencies are also determined in grapheme groups <G>, as are the predecessor and successor frequencies which reflect the dependencies of the assignment to phonemes of the preceding and succeeding graphemes.

[0032] Position-dependent frequency H^{pos} is understood as the frequency with which the grapheme at a specific position within a grapheme group <G> is assigned to a phoneme. Thus, for example, in the assignment of the grapheme group <ch> to the phoneme [C], the grapheme <c> is located at the first position, and the grapheme <h> at the second one. In this case, [C] is the voiceless palatal fricative or "Ich" sound, as in <Sicht>.

The frequency Hpos is calculated from

$$H^{pos}(g \rightarrow p|g \text{ in } < G > \text{ at Pos } i) = \frac{Z(g \rightarrow p|g \text{ in } < G > \text{ at Pos } i)}{N(p)}$$

[0033] The transitional frequencies are initialized by using the entries in a lexicon with words and their phonetic transcription, in the case of which the number of the graphemes coincides with the number of the phonemes. It is assumed that each grapheme is assigned to the corresponding phoneme. This is illustrated in Fig. 2 by the diagonally extending line.

[0034] This direct assignment is not always correct, as is shown, for example, by the example of <textlich> from Fig. 3, in which the line for the assignments does not extend simply diagonally. The number of the graphemes in the word <textlich> coincides with the number of the phonemes. There are 8 in each case. However, the letter <x> is mapped onto two phonemes [ks], and the letter group <ch> is mapped onto only one phoneme [C]. since such exceptions occur relatively seldom, however, they are of a correspondingly low weighting in the application of the relative frequencies. Moreover, all the frequencies which undershoot a specific threshold value are removed in a later correction step.

[0035] The assignments are counted, and the relative frequencies or transitional frequencies are determined from them.

[0036] The relative frequencies or transitional frequencies obtained in the preceding step are used to set up a matrix with transitional frequencies for each word in the lexicon, as is shown in Fig. 4 for the word <können>.

[0037] Four entries are permanently prescribed in this case. The entries at bottom left and top right must always be traversed, since they are the starting point and finishing point , respectively. They are therefore set to 1. By contrast, the fields at top left and bottom right can never be traversed. They are therefore set to 0. All other fields contain the corresponding transitional frequencies H(g->p).

[0038] In this initial assignment, <n> is assigned to the phoneme [9] (rounded half-open front vowel "ö"). Consequently 0.013 is set instead of numeral 0 in the corresponding fields. However, it may be seen that this frequency is much lower than the remaining frequencies. It is therefore of virtually no importance.

[0039] The individual matrix entries are now multiplied in each case by the maximum of the adjacent entries in order to calculate the path. Since only the movements upward, to the right or upward to the right are permitted, only the values on the left, at the bottom and at bottom left starting from the respective matrix entry are considered for determining the maximum.

[0040] If during the determination of the maximum value the matrix entry at bottom left (diagonally) starting from the respective matrix entry and one of the other two entries are of equal magnitude, the diagonally situated matrix entry is regarded as maximal.

[0041] The multiplication begins with the first entry at bottom left, use being made in the determination of the maximum values of the modified entries of the matrix respectively resulting from the multiplications.

[0042] The first column and the lowermost row represent special cases, since there is no left-hand or lower neighbor. Here, the current entry is always multiplied by the lower or left-hand entry. The individual products resulting are illustrated in Fig. 5.

[0043] The accumulated frequency at the final point at top right is therefore the product of the entries or frequencies on the optimal path from the starting point to the finishing point.

[0044] A step direction from matrix entry to matrix entry is determined by determining which of the three preceding matrix entries was maximal. Starting from the matrix entry for the last phoneme and the last grapheme (top right), a path is respectively defined through the matrix along the determined step direction up to the matrix entry at bottom left. The

matrix elements belonging to the path define the assignment of graphemes to phonemes of the word.

[0045] Subsequently, post-treatment is carried out for further improvement. The post-treatment serves to check the decisions made, taking account of the grapheme context and phoneme context.

[0046] Firstly, after execution of the described assignment of graphemes to phonemes for each word of the lexicon, these assignments are used to determine the relative frequency with which a phoneme is produced by two or more graphemes, or two or more phonemes are produced by a grapheme, that is to say the position-dependent frequency Hpos.

[0047] Subsequently, the assignment of graphemes to phonemes within a word is corrected with the aid of the position-dependent frequencies. Consideration is given for this purpose to Fig. 6A which corresponds in structure to Fig. 5. The previously described method supplies, for example, for the English word <yield>, the assignment

yi e l d

to j i: l d

since the frequency of the assignment of the grapheme <i> to the phoneme [j] is higher (here 0.04) than the frequency of the assignment to the phoneme [i:] (here 0.03).

[0048] The position-dependent frequencies show, however, that the frequency of the assignment of <i> to the phoneme [j] is low when <i> is located at the second position of the grapheme group <yi>. By contrast, the frequency of the assignment of <i> to the phoneme [i:] is high when <i> is located at the first position of the grapheme group <ie>.

[0049] This corrected assignment is also supported by the consideration of the position-dependent frequency of <e>. The frequency of the assignment of <e> to the phoneme [i:] is low when <e> is located in front of <l>. By contrast, the frequency of the assignment of <e> to the phoneme [i:] is high when <e> is located at the second position of the grapheme group <ie>.

[0050] The assignment can therefore be corrected in accordance with Fig. 6B.

[0051] After execution of the corrected assignment for each word of the lexicon, these corrected assignments are used to determine the transitional frequencies and the position-dependent frequencies. These are used in further assignments.

[0052] In order to determine the relative frequencies, only those assignments are taken into account in which the matrix entry for the last phoneme and the last grapheme (top right)

overshoots a prescribed threshold value after execution of the multiplications outlined. This matrix entry corresponds to the product of the transitional frequencies along the best path. The magnitude of this product is therefore used as a criterion as to whether this path is to be accepted or not.

[0053] The method is executed in several iterations. In this case, the threshold value is high at the start and is reduced after each iteration. Consequently, at the start only those assignments are accepted which are correct with relative certainty. Since all frequencies are less than 1, the length of the word also enters indirectly into the product. The more factors the product has, the smaller it becomes. Thus, at the start it is predominantly the assignments of short words that are accepted. With short words, the probability of finding a wrong assignment is smaller than in the case of long ones.

[0054] The assignments in the case of which the product of the transitional frequencies has overshot the threshold value are used to obtain the new statistics. Even in the case of the first evaluation of the statistics thus obtained, most of the errors which have resulted from the one-to-one initialization of the frequencies have vanished. Moreover, it is also checked how frequently each grapheme-phoneme assignment has occurred. If the ratio undershoots a threshold value, this assignment is ignored, and thus not further used when the matrices are next filled up.

[0055] The result is an assignment of the graphemes to the phonemes for the entire lexicon. Furthermore, a list is obtained showing which phoneme or which phoneme group can be produced by which graphemes, for example [tS] in English by <ch>, <cz>, <c>, <tch>, <cc>, <t> and <che>.

[0056] The invention has been described in detail with particular reference to preferred embodiments thereof and examples, but it will be understood that variations and modifications can be effected within the spirit and scope of the invention.